Inhibitory control and the speech patterns of second language users

[1] Malgorzata Korko* and [2] Simon A. Williams

[1] Department of Psychology, Middlesex University, The Burroughs, London, NW4 4BT, United Kingdom

[2]Sussex Centre for Language Studies, University of Sussex, Falmer, Brighton, BN1 9SH, United Kingdom

Word count (exc. figures/tables): 10,326

*Requests for reprints should be addressed to Malgorzata Korko, Department of Psychology, Middlesex University, The Burroughs, London, NW4 4BT,United Kingdom (e-mail: M.Korko@mdx.ac.uk).

1. Introduction

Production of speech is a complex cognitive phenomenon. The multiple processes involved require: the planning of what to say, retrieval of relevant material from the mental lexicon, construction of syntactic frames, encoding of the formulated message into a phonological structure and the execution of the appropriate articulatory gestures to obtain the desired speech signal (Dell, 1986; Garrett, 1980; Levelt, 1989). Second language (L2) users and bilingual speakers are faced with the additional challenge of having to resolve competition that may arise from simultaneous activation of representations from across the two languages (Kroll, Bobb, Misra, & Guo, 2008; Colomé & Miozzo, 2010; De Groot, 2011). Such co-activation and the need to eliminate non-salient information may occur at any stage of the language production process. That more than one language is active and competes for selection, despite the speaker's intention to use only the target language, has been demonstrated in observational studies (intrusions from a non-target language being interpreted as temporary control failure, e.g. 'slips of the tongue' in Poulisse & Bongaerts, 1994; Poulisse, 1999; interlingual blends e.g. "Springling" as a combination of "spring" in English and "Frühling" in German in Green, 1986); and with the use of experimental paradigms (e.g. cognate facilitation in Costa, Caramazza & Sebastian-Galles, 2000; picture-word interference in Hermans, Bongaerts, De Bot & Schreuder, 1998; tip-of-the-tongue (TOT) in Gollan & Acenas, 2004; interlingual homographs in Martin, Macizo & Bajo, 2010, and gender congruency in Morales, Paolieri & Bajo, 2011). The presence of more than one language can interfere with production at various levels. Within the studies on bilingual language processing in which concurrent activation of two languages has been induced experimentally, interference has been shown to occur at the levels of phonological (e.g.

cognate naming in Costa et al., 2000; phoneme monitoring in Colomé, 2001), semantic (e.g. interlingual homographs in Martin et al., 2010) and syntactic encoding (e.g. gender congruency in Morales et al., 2011).

Given the competitive nature of bilingual language production, when alternative solutions are available in both L1 and L2, or when a representation in the non-target language is more readily accessible due to a higher level of activation, how is the target word selected? Most answers to this question assume some form of cognitive control, although few converge on the nature and locus of such control. Proponents of the language-specific view on lexical access (Costa, 2005; Costa & Caramazza, 1999; Costa, Miozzo & Caramazza, 1999) maintain that despite concurrent activation of representations in both languages, bilingual speakers are able to direct their attention to alternatives in the intended language as these do not enter into competition with candidates in the language currently not in use. According to this model, access to the relevant lexical unit is restricted in an *a priori* fashion - based on the language cue, the speaker actively selects the target lemma, without the need to rely on inhibitory processes. In contrast, accounts of language-non-specific lexical selection, of which Green's (1998) inhibitory control model (ICM) is probably the most popular, postulate an *a posteriori* mechanism that allows for the suppression of non-intended representations once these have been activated by their corresponding lexical concepts (the reactive nature of inhibition) and in a magnitude that is proportional to the level of their activation (the more strongly activated the representation, the more inhibition is needed). This is not to say that active selection and inhibition are the only candidates to aid the reduction of cross-language interference. Other cognitive processes, such as working memory, mental flexibility, goal updating, planning and self-monitoring, collectively referred to as executive functions or executive control, are likely to support bilingual production; however, to date, inhibitory mechanisms have attracted most interest, occupying centre stage in behavioural (Festman & Münte, 2012; Hermans et al. , 1998; Kroll, Bobb & Wodniecka, 2006; Levy, McVeigh, Marful, and Anderson, 2007; Lee & Williams, 2001; Linck, Hoshino & Kroll, 2008; Linck, Kroll & Sunderman, 2009; Martin et al., 2010; Morales et al., 2011; Meuter & Allport, 1999; Poulisse, 1999; Pivneva, Palmer & Titone, 2012), neuroimaging (Abutalebi et al., 2008; Abutalebi & Green, 2007), and electrophysiological research on cross-language interference (Christoffels, Firk & Schiller, 2007; Misra, Guo, Bobb & Kroll, 2012; van Asche, Duyck & Gollan, 2013).

Meuter and Allport (1999) were among the first to confirm the claim that inhibitory control may be a mechanism that supports lexical selection. An important observation made about speakers prompted to switch from one language code to another was that participants took significantly longer to name the presented stimuli (Arabic numerals) in the switching condition (when switching from L1 to L2 and vice versa) than in the non-switching condition. Crucially, greater cost (i.e. longer naming latencies) was associated with switching from the less dominant L2 into the more dominant L1. Meuter and Allport (1999) explained this altogether counterintuitive result with the need of bilingual speakers to overcome the residual inhibition of L1. In other words, switching into L1 was more difficult because it was suppressed during the preceding trial in which L2 was produced (but see Costa and Santesteban (2004) for an alternative interpretation based on the lack of asymmetrical switching costs in highly proficient bilinguals). More recently, inhibitory control mechanisms have been implicated in the resolution of competition arising from the co-activation of phonological, semantic and syntactic representations across two different languages. Not only did the bilingual speakers in Martin et al.'s (2010) and Morales et al.'s (2011) studies take longer to respond to interlingual homographs and nouns of different grammatical gender across the L1 and L2 respectively; they also took longer to process subsequent tasks requiring renewed access to previously ignored information. Such a reduction in performance was associated with a demand to override inhibition that was evidently applied to previously non-intended, but activated representations. Thus, Martin et al.'s (2010) participants were slower to respond to interlingual homographs whose meanings were previously 'deactivated' on the relatedness judgement task. Similarly, Morales et al. (2011) found that Italian-Spanish bilinguals were less efficient at retrieving definite articles in their L1 if those were previously suppressed on gender incongruent trials.

In the majority of studies, speed of processing has been identified as the primary cost associated with cross-language competition, and the potential use of inhibitory control as the means to resolve such competition. Most such studies have explored experimental contexts in which single word production is required, and only a handful of authors have investigated the involvement of inhibitory control in contexts where multi-word utterances or prompted speech has been generated (including Engelhardt, Nigg & Ferreira, 2013; Festman, 2012; Pivneva et al., 2012). Festman (2012), for example, demonstrated that in speech elicited during a bilingual interview, bilinguals with lower executive functions ("switchers") produced more errors of CLI than "non-switchers"; however, the two groups were not significantly different from each other in other aspects of speech, such as fluency, syntactic complexity, grammatical correctness and word finding difficulties. It is debatable, however, to what extent the Aachener Aphasie Test (AAT, Huber, Poeck, Weniger & Willmes, 1983) adopted by Festman (2012) to assess language proficiency could adequately measure the various parameters of bilingual speech. The ATT was originally designed to qualitatively evaluate spontaneous speech in aphasic patients, which puts into question whether the 5-point scale used by the four judges in Festman (2012) was sensitive enough to detect differences in the features of interest in the speech produced by bilingual speakers with no neurological problems.

To investigate the relationship between aspects of L1 fluency (viz. filled pauses, unfilled pauses, repetitions and repairs[1]) on the one hand, and individual differences in intelligence (e.g. processing speed) and executive function (including inhibitory control as measured with the Stroop and stop-signal tasks) on the other, Engelhardt et al. (2013) elicited and quantified speech from a sample of 106 adolescents and adults by employing a sentence production task. The participants had to generate sentences based on pictures depicting both animate (e.g. *girl*) and inanimate (e.g. *bicycle*) objects and a verb (in either an unambiguous past participle form e.g. *ridden* or a form that could be used as past tense or past participle e.g. *moved*). On half of the trials animate objects were presented first, followed by inanimate objects. When presented in this order, with an unambiguous past participle verb (*ridden*), the animate object primes the speaker towards an active grammatical construction, yet the verb that has to be embedded in the sentence necessitates the use of the passive voice. Such a configuration introduces conflict, which needs to be resolved before a sentence is articulated. Engelhardt et al. (2013) hypothesised that individuals with poorer inhibitory control may begin to speak before they plan their utterance in full, which may result in disfluencies - an assumption that found some support

[1] filled pauses: fillers such as *uh* and *um*; unfilled pauses: silent pauses; repetitions: unintended repeats of a word or a string of words (e.g. *the papaya… the papaya was sweet*); repairs: stopping an utterance and starting with a new word or phrase (e.g. *the mango… papaya*) (Engelhardt et al., 2013).

in the data. Approximately one-third of the variance in repair disfluencies (for example, when the speaker reversed grammatical roles in mid-sentence by saying “the girl” and then switching to “the bicycle”) was accounted for by individual variation in inhibitory control.

The failure to obtain a significant association between inhibitory control and other types of disfluencies (repetitions, filled and unfilled pauses) in Engelhardt et al. ’s (2013) study can be explained by the choice of the task, which may not have allowed for the production of a sufficient number and variety of disfluency markers. The design of the task could similarly have been prone to strategic effects, with participants potentially developing a strategy to wait until both objects and the verb had been presented before attempting to plan their utterance. In addition, the cut-off point adopted for unfilled pauses (1 second) may have further restricted the scope of the analysis, with most authors accepting thresholds in the range of 0.25 to 0.4 seconds for extended speech (e.g. Towell, 1987; Raupach, 1980). It would be expected that in the case of single-utterance production, where the processing demand is lower than that in extended speech, silent pauses equal to 1 second or longer would be a relatively rare occurrence. Interestingly, the number of unfilled pauses recorded in Engelhardt’s et al. (2013) study was disproportionately high, far exceeding the frequency of other types of disfluency. This observation, combined with the finding that about one quarter of the variance in unfilled pauses could be accounted for by intelligence, may additionally point to the use of a strategy on the part of the speaker and provides further motivation for investigating similar disfluencies in extended speech. The results should also be interpreted with caution as the participants were originally recruited for the purpose of another study, and so their language background (whether the sample consisted of native English speakers, mono- or multilinguals) is not fully known.

An investigation by Pivneva et al. (2012) appears to offer a middle ground methodological solution to the studies by Festman (2012) and Engelhardt et al. (2013) in that it extends beyond single utterance production and employs, although not exclusively, a quantitative measure of speech analysis. The authors explored an association between inhibitory capacity (as measured with a battery of anti-saccade, non-linguistic Simon, non-linguistic Stroop, and number Stroop tasks) and the efficiency with which English-French (less balanced) and French-English (more balanced) bilinguals produced L1 and L2 speech. In a modified version of the Map Task (Anderson et al., 1991), participants described a route first to a “hypothetical” listener (producing a monologue) and then to a confederate (engaging in a dialogue). The speech samples were rated by two independent judges on a scale from 1 to 9 along the following dimensions: clarity of content, fluency (smoothness of speech, absence of interruptions, hesitation, self-repairs and changes in speech rate) and nativeness (the extent to which the speaker sounded native-like). In addition, the samples were analysed using an acoustic-temporal measure defined as a ratio between individual vocalization duration and its prior silent pause duration (VD/PPD), with higher ratios reflecting greater ease of speech production. The study found no statistically significant relationship between inhibitory capacity and the spoken L2 output as assessed by the raters. Crucially, however, there was a main effect of inhibitory capacity on the VD/PDD ratio in L2 speech; the poorer the inhibitory capacity, the smaller the VD/PPD ratios and the greater the effort to produce L2 speech. Despite detailed assessment criteria and an excellent inter-rater agreement on the global output measures, subjective evaluations may have lacked the sensitivity to detect individual variation in L2 speech patterns. It is not clear either whether the map task, which was based on routes with landmarks that had word labels ascribed to them, was sufficiently taxing for the speakers to experience interference and resort to inhibitory control.

The current study seeks to extend this work by exploring the relationship between inhibitory control and more spontaneous L2 spoken language production from the perspective of individual differences. On the one hand it builds on previous studies that have centred on single-word production, on the other it takes a slightly different approach to the handful of studies that have looked at single utterances or prompted speech. As such, it continues in the tradition of Festman et al. (2010), Engelhardt et al. (2013) and Pivneva et al. (2012) and is primarily concerned with the question of how individual variation in the ability to suppress irrelevant and conflicting information is expressed in spoken L2 output. The study differs from previous investigations in three major ways. First, it targets speakers of English as a foreign language rather than bilinguals who acquired an L2 at a relatively young age as in Festman (2012) or Pivneva et al. (2012). Based on Green's (1998) inhibitory control model, which stipulates that the amount of inhibition needed to suppress a non-target language is relative to the proficiency of that language, these kinds of unbalanced bilinguals may need to exercise more control to suppress their dominant language. Second, it extends the scope of the analysis beyond the performance of tasks requiring the production of single words and utterances and focuses instead on the dynamics of prompted extended speech. It presents participants with a standardised written prompt to generate two minutes of extended speech, much like the so-called long turn in some public speaking examinations such as the UCLES FCE, CAE, CPE and IELTS[2]. Such monologic turns offer the advantage of producing a stretch of uninterrupted speech for analysis, with a better claim to 'verbal

[2] University of Cambridge Local Examinations Syndicate (UCLES) Cambridge English: First (FCE), Cambridge English: Advanced (CAE), Cambridge English: Proficiency (CPE) and International English Language Testing System (IELTS).

fluency' than the standard word production or sentence production tasks. An additional advantage of the task used in the current study, compared to sentence elicitation or route description, is that it puts increased processing demand on the speaker. It stands to reason that the greater the processing load, the greater the scope for the use of cognitive control.

Third, while the investigation by Festman (2012) provides first instances to the relation between prompted speech and inhibitory control, it relies exclusively on global output measures (viz. the ATT test) to evaluate oral L2 performance. This study adopts a more fine-grain approach to the analysis of verbal output, extending the repertoire of quantitative measures used in Engelhardt et al. (2013) and Pivneva et al. (2012). To provide a detailed, objective assessment of spoken L2 performance, we analysed speech along the following dimensions: filled pauses[3], frequency and duration of silent pauses[4], repetitions[5], reformulations[6], articulation rate[7], total number of words and pruned words[8] produced, and performance errors[9] (see Table 2 for extended definitions and examples). In addition, a distinction was made between silent pauses at mid-clause and end-clause position in the utterance, with clause-internal pauses perceived as more disruptive than pauses at clause boundaries - a feature of natural prosody (Pawley and Syder, 2000). The choice of these variables was motivated by the claim that while some disfluency markers are used as signalling devices (Corley & Stewart, 2008), most disfluencies arise from processing

[3] voiced hesitations, such as *um* and *uh*, sometimes called 'fillers' (Corley & Stewart, 2008).
[4] an unvoiced delay, a temporary suspension of speech activity (Clark 2006: 244)
[5] consecutive, and semantically redundant, production of the same phoneme, syllable, word or phrase.
[6] instances in which the speaker abandons an original utterance/word and starts with a different one.
[7] the speed of a speaker's delivery measured in words per minute
[8] the total number of words disregarding filled pauses, repetitions and reformulations (cf pruned syllables in Bosker et al., 2012).
[9] errors resulting from attentional lapses or failures of inhibition as opposed to 'proficiency errors' resulting from gaps in the speakers' linguistic knowledge.

difficulties (Levelt, 1989; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). It is hard to identify the underlying cause of these difficulties, but it has been suggested that they may involve planning, monitoring, retrieval and/or flexibility problems (Clark & Wasow, 1998; Clark, 2006). In unbalanced bilinguals, in whom L1 is the more dominant language, some of these problems may be traced back to cross-language interference and the efficiency with which speakers select the intended representations in the face of intrusions from the non-intended language.

Following this line of argument and the previous work on the relation between executive function (specifically on inhibitory control) and spoken language production beyond a single word utterance, we hypothesised that the L2 speech produced by individuals with poorer inhibitory control will be characterised by reduced fluency. This could manifest itself in both increased frequency and prolonged duration of silent pauses, particularly mid-clause silent pauses, and decreased articulation rate – aspects of speech which may not only signal hesitations in the speaker's planning process, but also indicate transient difficulty with lexical access when there are more rivals for selection (Goldman-Eisler, 1961). As repetitions are considered a stalling tactic, reflecting the speaker's attempt to compensate for a cognitive difficulty, such as retrieval of an upcoming word. (e.g. Clark & Wasow, 1998; Dörnyei and Kormos, 1998), instances of repetition were expected to be higher among those with less efficient inhibitory mechanisms (but see Levelt (1983) and Tannenbaum, Williams & Hillier (1965) for an alternative interpretation of repetitions). Similarly, frequent backtracking, both in the form of self-initiated repairs and false starts, could relate to difficulties in resolving cross-language competition. The speaker is unable to quickly 'deactivate' a non-intended representation, so the speech production process comes to a halt and the original utterance or part of it must be abandoned. The frequency of filled pauses in prompted extended speech, on the other hand, was thought to bear little or no relation to inhibitory processes as this type of pausing phenomena has been documented to be language-specific, primarily serving a signalling function (O'Connell & Kowal, 1972; O'Connell, Kowal,& Hőrmann, 1969), although the literature is far from clear on this issue, with some authors (e.g. Bortfeld, Leon, Bloom, Schober & Brennan, 2001; Clark and Fox Tree, 2002) relating such interruptions to planning difficulties. The study also looked at the relationship between inhibitory control and the frequency of performance errors. Such errors, produced but subsequently recognised as errors by the participants in a post-hoc error identification task, were expected to correlate negatively with inhibitory control.

Is the speech of individuals with poor inhibitory control more hesitant? Is it characterised by increased pausing or frequent self-corrections? Are L2 learners who are by nature more resistant to interference on average slower speakers? Do such learners display greater susceptibility to performance errors? To address these and similar types of question, each of the examined speech variables was correlated with inhibitory control, while controlling for age and L2 proficiency – factors known to affect spoken language production (e.g. Bortfeld et al., 2001; Horton, Spieler & Shriberg, 2010; Kormos and Denes, 2004). The central aim of the current paper was therefore to examine the extent to which inhibitory control accounts for individual variation in L2 speech production, above and beyond age and L2 proficiency level, with special emphasis on the fluidity of the speech.

2. Method

2.1 Participants

Eighty-two students with English as their L2 took part in the study ($N_{females}$=47, M_{age}=26.35, SD_{age}=6.49, 19-46 years). The students were recruited from a British university, where they were attending general EFL (English as a foreign language) classes at intermediate to advanced level for 21 hours a week. They reported as their dominant languages Chinese (n=28), Arabic (n=8), Thai (n=7), Spanish (n=7), Turkish (n=5), Japanese (n=5) and 13 others (see Appendix A for all the reported first languages and their frequencies).

Each speaker's L2 proficiency level was formally assessed by two experienced EFL teachers, who independently rated each participant's speech sample post hoc by listening to the recording and applying the appropriate descriptors from the public version of the IELTS Speaking Band Descriptors (IELTS, n.d.). The raters awarded scores on the scale from 0 to 9 for four criterion areas: Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy, and Pronunciation. As the inter-rater reliability analysis revealed an acceptable level of agreement for all the descriptor types, the raters' scores were averaged to produce a mean band score for each criterion area. Table 1 reports the means, standard deviations, ranges and Intraclass Correlation Co-efficients (ICCs) for the mean Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy and Pronunciation band scores, and the overall spoken L2 proficiency score, which is based on the mean of the four aggregated band scores. Participants' mean speaking band scores ranged from 4.5 to 9, corresponding to levels B1 (independent user) to C2 (proficient user) according to the Common European Framework of Reference (Council of Europe, 2001).

Table 1 Descriptive statistics and Inter-rater reliability for the overall and composite L2 spoken proficiency scores

2.2 Materials and Procedure

Participants were tested individually in a quiet room. Upon signing the consent form and completing a short demographic and language background questionnaire, their focal colour recognition was assessed using four colour patches (blue, green, red and yellow) presented on the computer screen. Participants subsequently performed two computerised inhibitory control (IC) tasks: Stroop and shape matching. After that, they were given a speech production task with the aim of eliciting 2 minutes of uninterrupted L2 speech, followed by an error identification task. Participant responses on both the speech production and the error identification tasks were audio-recorded for further transcription and analysis. The whole testing session lasted approximately 30 minutes.

Stroop task

To assess individual differences in the ability to suppress prepotent, automatic responses, we administered a modified, computersied version of the Colour-Word Stroop Task (Stroop, 1935). PsychoPy (Peirce, 2006) was used to present the stimuli and collect response data. In the modified version of the Colour-Word Stroop Task, participants were instructed to select the colour of the stimulus (a colour word presented in lower case font against black background) as quickly and as accurately as possible, while ignoring its name. They were asked to respond manually by pressing a corresponding button on the keyboard: B for blue, G for green, R for red, and Y for yellow[10]. Participants were given a brief practice session before the actual task so they could familiarise themselves with the procedure and the position of the four keys on the keyboard. There were two experimental conditions: congruent and incongruent[11]. The congruent condition consisted of word stimuli that were presented in the same ink as the colour name (e.g. the word 'blue' in blue ink). In the incongruent condition, the words were presented in a different ink (e.g. the word 'blue' presented in red ink). There were 48 trials, half congruent, half incongruent. The trials were presented in a randomised order. Each trial began with a fixation cross (500 ms), followed by a blank screen (300 ms), and then the word stimulus appeared for 2,500 ms or until a response was made. A blank screen was presented following each trial at an interstimulus interval varying from 1000 to 1500 ms. It was not possible to backtrack if an erroneous response was detected.

Inhibitory control on the Stroop task was understood in terms of an interference effect which was obtained by subtracting mean reactions times (RTs) and mean error rates (ERs) on the congruent trials from mean RTs and mean ERs on incongruent trials. Only the correct trials were included in the analysis of reaction times (RTs). Based on the outlier labelling rule, with g = 2.2, which was applied to screen for outliers among RTs and ERs (Hoaglin &

[10] Although the interference effect on the Stroop task is typically more pronounced when vocal responses are required, the effect has also been established in previous manual Stroop studies, e.g. Besner, Stolz & Boutilier, 1997; Coderre and van Heuven, 2014; Heidlmayr et al., 2014; Kousaie and Phillips, 2012; a significant interference effect was also obtained in the current study, see Table 4).

[11] We decided not to include a baseline condition for the lack of consensus on which types of neutral stimuli are most appropriate (McNamara, 2005). In the case of repeated exposure to the 'neutral' stimuli, such as a string of symbols (e.g. XXXX) or the word 'blank/neutral', participants may habituate to such stimuli, which decreases the processing demand on neutral trials and artificially inflates the benefits associated with performance on critical trials (Jonides and Mack, 1984). The processing complexity of pronounceable non-words, on the other hand, is higher than that of colour words, which may, in turn, artificially increase the response latencies on neutral trials (McNamara, 2005).

Iglewicz, 1987), an overall of 2.4% of the Stroop Task data points were excluded from the analysis due to extreme values.

Responding correctly in the incongruent condition requires participants to resolve the conflict between the well-learned reading response and the colour-naming response (MacLeod, 1991). As participants must engage cognitive control on incongruent trials to inhibit an automatic response (word reading), their performance is slowed and their accuracy diminishes relative to congruent trials. Longer reaction times and higher error rates are therefore associated with poorer inhibitory control.

Shape matching task

An abridged, computerised version of the shape matching Task (DeSchepper & Treisman, 1996) designed in PsychoPy (Peirce, 2006) was used in the study to evaluate an ability to resist distractor interference. Friedman and Miyake (2004) list the shape matching task among the standardised measures that assess 'the ability to resist or resolve interference from information in the external environment that is irrelevant to the task at hand' (Friedman & Miyake, 2004: 104). The selection of a non-linguistic inhibitory control task for the purpose of the current analysis was motivated by the claim that language control should not be conceived of as part of the language system per se, but rather as the result of a domain-general executive function (e.g. Costa, 2005; Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000).

In the modified version of the shape matching task, participants were presented with abstract shapes on the screen: a green target shape, which was either presented alone (no distractor condition) or was superimposed on a red distractor shape on the left side of the fixation point (distractor condition), and a white shape that appeared alone on the right side of the fixation point (Fig. 1). Participants were asked to manually indicate by pressing the corresponding button on the keyboard (specially labelled keys: 'Y' for 'yes' and 'N' for 'no') whether the green target shape on the left matched the white shape on the right, ignoring the distractor shape when one was present. Before the actual task, the participants received a short practice session to familiarise themselves with the procedure and the response-key mappings.

Figure 1 Shape Matching Task. Participants indicated whether the green target shape on the left matched the white shape on the right, ignoring the red distractor shape when one was present.

The experiment consisted of 48 trials, half of which were distractor trials; the other half contained no distractor. The trials were presented in a randomised order. In each trial, a fixation point appeared for 500 ms, followed by a blank screen (300 ms) and the shape stimuli. The latter were displayed for a maximum duration of 3000 ms or until a response was made. A blank screen was presented following each trial at an interstimulus interval varying from 1000 to 1500 ms.

The interference effect on the shape matching task was defined as the difference between the mean RTs on distractor trials and the mean RTs on no-distractor trials. Only the correct trials were included in these averages. The same difference was computed for ERs. The outlier labelling rule, with $g = 2.2$, was applied to screen for outliers among RTs and ERs

(Hoaglin & Iglewicz, 1987). Overall, 1.2% of data points were excluded from the analysis due to extreme values.

On distractor trials participants are expected to suppress a visual stimulus that interferes with the recognition process (deciding whether the target shapes are the same or not). Therefore, the quicker participants can filter out such irrelevant information to decide whether the target stimuli are the same or different, the more efficient their inhibitory capacity. In terms of accuracy, higher ERs on the shape matching task reflect poorer inhibitory control.

Speech production task

Following the inhibitory control tasks, participants completed the speech production task. Each participant was given a topic with a semi-structured prompt and asked by an English native speaker to speak to it uninterrupted for two minutes. The topics comprised a selection adapted from Hashemi and Thomas (2011) and from Allen et al. (2007), e.g. ‘Describe a journey you remember well’ (see example in Appendix B), presented on a cue card and assigned to each participant at random. Before starting to speak, the participant was given 1 minute to think about the topic and to make notes if they wished. A pen and paper were provided for this purpose.

Compared to previous tasks which served as the basis for a quantitative analysis of oral fluency (e.g. sentence generation in Engelhardt et al. 2013 and route description in Pivneva et al., 2012), the speech production task employed in this study enables the speaker to follow a suggested train of thought in their own way, imposing minimal lexical and syntactic structures via the very general guidance of the prompt. Much as this reduces experimental control over the speaker's output, the task has the advantage of providing scope to observe a greater number and variety of disfluency markers. The task also imposes additional processing demands beyond word or utterance production, which increases the chances of engaging cognitive control on the part of the speaker as not only words but multiple utterances need to be planned and arranged into a coherent whole.

Participant responses elicited by the speech production task were audio-recorded, transcribed orthographically and coded independently by two raters. Two participants were excluded from the analysis due to suspected stammer and an insufficiently long speech sample (less than 100 words). From the transcripts, the following speech parameters were identified and tallied: filled pauses, repetitions, reformulations, and rater-identified speech errors. Their definitions and instances are presented in Table 2. Any discrepancies in the number and type of the tallied variables were re-evaluated and resolved by the raters.

Table 2 Speech parameters identified and tallied by raters in the transcribed speech samples

In addition, a number of temporal measures were examined. These were: number and total duration of silent pauses, number and total duration of mid- and end-clause silent pauses, and articulation rate. Number of silent pauses and total duration of pausing time were quantified using a *TextGrid* Silences script in PRAAT (Boersma & Weenink, 2013). The data set in this study is similar to that of de Jong and Bosker (2013), who compared different pause thresholds in monologic speaking tasks against a measure of L2 proficiency and conclude, 'for the purpose of L2 research, the traditional cut-off point of 250 ms is a good choice' (de Jong and Bosker 2013: 20), and accordingly a cut-off threshold of 250 ms was adopted here.

The frequency and total duration of mid- and end-clause silent pauses were calculated manually by one of the authors using Audacity 2.0.6 sound editor. Software measurement was not an option in this case as it was important to distinguish between more natural, prosodic pauses at syntactic and semantic boundaries and pauses that are inserted between words or constituents. One by one, MPEG Layer-3 audio files for each participant were uploaded to Audacity and the wave-form maximised. The counter was set to 'length' and silent pauses highlighted and finely adjusted before the total length was read off and noted both in the transcript and in a separate Excel wordsheet. In this way, the silent pauses of the 82 participants were identified, timed and recorded qualitatively on the transcripts. Next, the pauses on the transcript were coded according to their position in T-units: 'one main clause and all its attendant subordinate clauses and nonclausal units' (Lennon 1990: 406), i.e. whether they marked the end of a clause or appeared mid-clause. For example,

it was not a good hotel because it had only three [0.250 mid] stars but for me it was [0.495 mid] admirable [1.071 end]

Ten randomly selected speech samples were additionally analysed by two independent raters for the numbers of mid-clause (non-juncture) and end-clause (natural, juncture) pauses. Articulation rate was expressed as the number of pruned words (with the exclusion of repetitions and filled pauses) produced in one minute.

The transcripts were also examined by two independent raters for errors (rater-identified errors), indicated by non-standard grammatical and lexical forms. A speech segment was categorised as a lexical error if it had a non-standard form or meaning, e.g. *I travelled to Rhodos in **Greek*** [Greece] (form), or *my friend kept on playing and he **gained** … eight hundred Euro* (meaning). It was categorised as a grammatical error if it had a non-standard syntactic form, e.g. *He's living in America since fifty years I think* (two errors: inappropriate use of the present continuous for the present perfect continuous; and misuse of preposition 'since' for 'for').

To assess inter-rater reliability, Intraclass Correlation Co-efficients (ICC) were computed for all the coded speech parameters. An inter-rater reliability across all of these variables as measured with a two-way, mixed model absolute agreement test was within an acceptable range of agreement, with ICC coefficients, means, standard deviations and ranges for all these measures reported in Table 3.

Table 3 Descriptive statistics and inter-rater reliability for L2 spoken output measure

Error identification task

Following the speech production task, participants performed the error identification task. Its aim was to obtain the number of performance errors (speech errors that result from temporary failure of cognitive control) as opposed to proficiency errors (errors that are attributed to the speaker's linguistic competence). The procedure of the error identification task follows Kormos (2000). In her study, participants commented on their spoken performance after listening to a recording of their speech. They were asked to stop the playback whenever they noticed breakdowns or self-repairs and provide a gloss or comment (Kormos 2000: 352). In this study, participants were required to orally identify as many mistakes as they could while listening to their own audio-files. The process was recorded and any comments subsequently transcribed. To allow for the fact that spoken data is transitory, and its detail is therefore more difficult to attend to, participants were invited to listen to the recording twice.

Grammar and lexical errors identified by the participants were tallied and subtracted from the total number of speech errors identified in the transcripts by two independent raters (rater-identified errors). This allowed us to obtain two measures of spoken L2 output: performance errors (errors identified by the participants) and proficiency errors (errors identified by the two raters minus the performance errors identified by the participants).

3. Results

Results for individual inhibitory control measures are reported in the first instance, followed by correlations between the variables of interest and regression analyses to establish the impact of the predictor variables (inhibitory control as measured with the Stroop and shape matching tasks) on oral L2 performance (individual aspects of L2 speech) above and beyond age and L2 proficiency level.

Inhibitory control measures

The Stroop interference effect was significant both for reaction times (RTs), $t(79)=16.06$, $p<.001$, $\eta_p^2=.765$, BCa 95%CI [183, 235] and error rates (ERs), $t(79)=6.83$, $p<.001$, $\eta_p^2=.371$, BCa 95% CI [3.2,5.8]. The interference effect observed on the shape matching task was also significant for both the RTs, $t(80) = 12.5$, BCa 95%CI [177, 243], $p < .001$, $\eta_p^2=.670$, and the ERs, $t(80)=2.76$, BCa 95% CI [.51,3.1], $p <.01$, $\eta_p^2=0.087$. Means and standard deviations for both the latencies and error rates across the two types of trials on the Stroop and the shape matching tasks are reported in Table 4.

Table 4 Means (M) and standard deviations (SD) for the interference effect and the two types of trials in the Stroop and the Shape Matching Tasks expressed in reaction time (RT) and error rate (ER)

Intercorrelations among and between age, L2 proficiency, IC and L2 spoken output measures

Tables 5 and 6 provide Pearson's correlation co-efficients for zero-order and partial correlations respectively, describing the relations between the variables of interest.

While there was a trend towards a positive correlation between age and overall spoken L2 proficiency (r =.212, p=.056), only the association between age and lexical resource reached statistical significance (r=.279, p=.011). Lexical resource as one of the global output measures used to assess L2 proficiency in the current study refers to the range of lexis, its accuracy and appropriacy (IELTS, n.d.). Therefore, as revealed in the present analysis, the selection of lexis appeared to become more skilful and diverse with the speaker's age - a finding consistent with Horton et al. (2010). Quantitatively measured L2 speech parameters showed a negative correlation between age and the frequency and duration of silent pauses (treated collectively) and the frequency and duration of mid-clause silent pauses (with rs ranging from -.247 to -.281, p<.05), indicating less frequent and shorter pausing with increasing age. After controlling for L2 proficiency, there was also a significant positive correlation between age and filled pauses (r=.254, p=.031), which is in line with Bortfeld et al. (2001) and Horton et al. (2010), who reported a general increase in the use of fillers (e.g. "uh" and "um") with age. While no other statistically significant associations were found between age and the remaining L2 speech parameters, these findings alone further justify the claim to partial out the effect of age when looking at inhibitory control as a predictor of spoken L2 performance.

There was a moderate positive correlation between age and Stroop interference (measured in RT) on the one hand, with r =.387, p<.01, and age and shape matching interference (measured both in RT and ER), on the other, with r=.321, p<.01 and r=.247, p=.025 respectively. In other words, greater interference effects were observed with increasing age, which is in line with previous findings reporting an age-related decline in inhibitory control function (e.g. Hasher, Stoltzfus, Zacks, & Rypma, 1991).

Significant correlations were observed between the overall L2 proficiency and the majority of the quantitatively measured L2 speech parameters, with the exception of the total silent pause duration and end-clause silent pause duration. This is understandable as pauses at clause and sentence boundaries are natural prosodic markers, which should have a negligible effect on the listener's perception of the speaker's intelligibility or proficiency level. The measures which appeared to correlate most with the raters' judgement of L2 proficiency were articulation rate, percentage of pruned words and percentage of proficiency errors. There was a tendency to speak at a faster rate (r=.658, p<.01), produce more pruned words (r=.617, p<.01) and make fewer proficiency errors (r=-.523, p<01) with increasing L2 proficiency level.

Importantly, among the analysed L2 spoken output measures, only reformulations and the total frequency and duration of silent pauses correlated significantly with the interference effect as obtained on the Stroop task and indexed with ER. Lower accuracy observed on the Stroop task was associated with a greater percentage of reformulations recorded in L2 speech (r=.227, p=.044). Higher ERs on the Stroop task were similarly linked to increased pausing, both in frequency (r=.264, p=.029) and duration (r=.231, p=.049). The strength of these relationships increased to r=.274, p=.021 (reformulations), r=.269, p=.023 (frequency of silent pauses) and r=.236, p=.047 (duration of silent pauses) respectively after partialling out the effects of age and L2 proficiency.

Table 5 Correlations between age, L2 proficiency, IC and L2 spoken output measures

Table 6 Partial correlations between measures after controlling for age and L2 proficiency

Regression Analyses

To determine the unique contribution of inhibitory control to spoken L2 performance after accounting for the effects of age and L2 proficiency, a series of hierarchical multiple regressions was carried out. With individual measures of L2 spoken output as dependent variables, age was included as a control variable in the first block and the overall L2 proficiency in the second, followed by individual measures of inhibitory control in the third block. The only L2 spoken output measures that could be reliably predicted by inhibitory control (expressed as the Stroop interference effect and indexed with ER), above and beyond age and L2 proficiency, were reformulations and the total frequency and duration of silent pauses. In the model explaining the unique contribution of inhibitory control to the occurrence of reformulations in L2 speech, age was a non-significant contributor, while the overall L2 proficiency accounted for ca. 17% of the variance in the reformulation rate, with a significant R^2 change of ca. 15% [$F(2,78)=7.65$, $p<.01$]. When inhibitory control (expressed as performance accuracy on the Stroop task) was taken into account, the whole model (age, L2 proficiency and inhibitory control) accounted for more than 22% of the variance in the percentage of reformulations observed in L2 speech. In other words, adding inhibitory control to the model increased its predictive capacity for the use of reformulations in a statistically significant way by ca. 5%, [$F(3,78)=7.07$, $p <.01$). The β co-efficients for the selected predictor variables in the final models are provided in Table 7. Age but not L2 proficiency was a significant predictor of the total frequency and duration of silent pauses, accounting for ca. 6% of the variance [$F(1,78)=4.45$, $p=.038$]. When inhibitory control was factored in, the whole model accounted for 13% of the variance in the frequency of silent pauses [F(3,78)=3.46, p=.021] and 12% of the variance in the duration of silent pauses [F(3,78)=3.09, p=.033], again increasing the predictive capacity of the model by about 6%.

Table 7 Linear models of variables predicting individual L2 spoken output measures

4. Discussion

The aim of the present study was to demonstrate how inhibitory control as one of the cognitive mechanisms that have been proposed to reduce cross-language interference relates to spoken L2 performance. Specifically, the analysis focused on establishing whether a general ability to suppress irrelevant information can predict the speed with which the non-dominant language is produced and the different types of disfluencies that occur in it.

Our hypothesis that the L2 speech produced by individuals with poorer inhibitory control would be generally less fluent was only partly confirmed. Inhibitory control (but only expressed as performance accuracy on the Stroop task) significantly predicted the occurrence of reformulations and the total frequency and duration of silent pauses in L2 speech, above and beyond the speaker's age and L2 proficiency level. Higher error rates on inhibitory control tasks indicate poorer inhibitory capacity. It follows that those individuals who are more prone to errors on such tasks, failing to resolve competition between conflicting responses and/or conflicting stimuli, are more likely to pause and reformulate the initiated utterances. This is partly in line with Engelhardt et al. (2013), who found that repair disfluencies in L1 speech were specifically related to individual differences in inhibitory control, and not to intelligence or mental set shifting.

To fully account for the link between inhibitory control and the tendency for speakers to reformulate their utterances, as corroborated in the present study, it is worth recalling the significance of this particular aspect of speech. Reformulations in this work were understood as false starts on the one hand, and as self-initiated repairs (Schegloff et al., 1977), on the other. While both entail an interruption to the flow of speech, in false starts, the speaker aborts an original utterance due to an unforeseen semantic, syntactic or phonological difficulty or an intrusive conceptual representation, and starts the utterance anew. In the case of a self-initiated repair, the speaker commits an error, stops, backtracks and corrects the deviant part of the utterance (Kormos, 1999; Levelt, 1983; Maclay and Osgood, 1959). False starts are thus thought to relate more to the lack of L2 competences; self-repairs, in turn, can be viewed as an attempt to rectify an accidental lapse (Kormos, 1999).

If reformulations are understood in the latter sense, as self-initiated corrections, it can be assumed that the less resistant the speaker is to unwanted information, the more errors will slip into his or her speech despite adequate L2 knowledge. Such errors, if intercepted by the self-monitoring system, are likely to be repaired. Overt repairs, in turn, contribute to a higher proportion of reformulations. Thus, based on the present findings, overt self-repairs may not only serve as evidence that some kind of meta-cognitive processes are in operation (Levelt, 1989; Postma, 2000), but also indicate insufficient inhibitory control mechanisms that fail to stop the activated but non-intended (conceptual, lexical, syntactic, phonological or articulatory) information in its track. It is less clear how individual variation in inhibitory control could account for reformulations understood in the former sense of the term, as false starts, where the main function of a reformulated message is to circumvent a conceptual, semantic, syntactic or phonological problem. The speaker lacks an adequate speech plan and so must suspend the utterance in midstream. Ideally, these two instances of reformulations should have been analysed as separate categories; however, due to the relatively low frequency of occurrence of this particular aspect of speech in the obtained samples, treating them collectively appeared to be the most logical solution.

Another type of dysfluency that was reliably predicted by inhibitory control (but, again, only in terms of performance accuracy on the Stroop task) was the frequency and total duration of silent pauses. The analysis revealed that poorer inhibitory control was associated with increased pausing, a finding which contradicts our original assumption, but which is partly consistent with Pivneva et al. (2012), where poorer inhibitory control was related to smaller VD/PPD ratios (smaller ratios reflect longer prior pause duration). It was expected that those individuals who inadvertently let non-intended representations enter their working memory, and consequently the flow of their speech, do so at the expense of planning, and as such their performance should be characterised by decreased pausing. However, while silent pauses can mark critical points in speech planning (e.g. Riggenbach, 1991) and be used more or less consciously as a rhetorical device to hold the floor or elicit a particular emotional response in listeners (e.g. O'Conell et al., 2010), they can also be interpreted as a sign of a production difficulty. The latter refers to the ease with which the speaker conceptualises what to say, selects a corresponding lexical representation, encodes it into a grammatical structure, and assembles and articulates its sounds. If spoken language production is understood as a competition-based process that involves activation of a wider set of representations, including ones from a non-target language, then inhibitory control could serve as a mechanism that narrows the focus of such activation and by doing so aids in the selection of the desired target. In the case of poorer inhibitory control, increased pausing may indeed reflect the time taken by the speaker to override the highly active but irrelevant representation from the language not currently in use. Pausing phenomena are thus delays in production, which may be a manifestation of conflict resolution attempts rather than of online planning.

Although statistically significant, these findings raise a number of important questions. First, why among a considerable number of spoken L2 output measures analysed in the present study did only the two speech parameters (rate of reformulations and silent pauses) correlate with the performance on inhibitory control tasks? The fact that the study produced insufficient evidence for the link between inhibitory control and the use of filled pauses is fairly easy to reconcile. It is plausible that the speakers mapped their L1 pausing patterns onto their L2 production. That filled pauses in speech are culture-specific has found support in a number of studies (e.g. Leal, 1995; Riazantseva, 2001; Riggenbach, 1991). Given the evidence and the relative heterogeneity of our sample's L1 background, with 19 different first languages reported, it is possible that the rate of filled pauses as reported in this study was skewed by hesitation patterns typical of participants' dominant language. It is also worth noting that filled pauses were inversely related to the duration of silent pauses ($r=-.270$, $p<.05$), pruned words produced ($r=-.351$, $p<.01$) and articulation rate ($r=-.357$, $p<.01$), suggesting that different mechanisms may be at play when such disfluencies arise.

It is less understandable why the remaining L2 speech variables analysed in the present study, namely repetitions, articulation rate, mid-clause silent pauses, and, most importantly, performance errors, did not correlate with either of the inhibitory control measures. It could be argued that various aspects of spoken L2 performance reflect the operation of different cognitive mechanisms. While reformulations may provide clues to the workings of inhibition, repetitions and articulation rate may be associated with distinct mental processes. Indeed, in Levelt's (1982) seminal study on self-repairs, editing terms, filled pauses and repetitions were taken as evidence of covert editing. Levelt (1983) concluded that what we say or intend to say is subjected to continuous mental scrutiny and that a corrective intervention on the part of the speaker, even before an error becomes apparent in the output, may interfere with the ongoing verbal performance, resulting in these disfluencies. This explanation is, however, difficult to reconcile with the present data, which point to a significant inter-correlation between reformulations and repetitions (r=.388, p<.01) and suggest that the two are related, possibly representing a common underlying mechanism. Future studies could use a latent variable approach to disentangle the potential contribution of a number of cognitive functions to the production of L2 speech, and demonstrate how this contribution translates into L2 speech patterns.

Second, based on the present findings, a question arose as to why, despite a marginally significant relationship between the Stroop and the shape matching interference effects (RTs) (r=.206, p=.067), suggesting shared variance between the two tasks, reformulations and silent pauses related solely to the performance on the Stroop task. A similar observation was, nota bene, made by Engelhardt et al. (2013), where self-repairs correlated with the Stroop task, but not with an alternative inhibitory control measure used in the study, the stop signal task. There are two potential explanations. While both tasks entail an element of suppression, they differ in the source of interference. The Stroop task is used to measure the efficiency with which the user overrides the dominant response (reading a word) and selects the required response (reporting the colour of the word). To perform the shape matching task, the user must maintain representations of two visual stimuli in active state, while ignoring a distracting visual stimulus. The Stroop task is thus more concerned with automaticity, where the nature of interference is predominantly motor - we have become so adept at reading that a string of letters invariably elicits an automatic reading response - and as such may bear more resemblance to speech production, where automatically activated but irrelevant representations or execution responses associated with such representations must be overridden. It is harder to draw a parallel between speech production and performance on the shape matching task, where interference has little to do with automaticity and is more perceptual in nature (Friedman & Miyake, 2004). Second, although the literature recognises the Stroop and the shape matching tasks as standardised measures of inhibitory control (e.g. Friedman & Miyake, 2004), the tasks may tap into processes other than inhibition. This was captured in the “unity/diversity framework” developed by Miyake et al. (2000), according to which tasks that are thought to involve executive functions share some commonalities, but differ on a number of dimensions. Because the tasks used in the present study are not pure measures of inhibition, even if a correlation is established it cannot be said with certainty that a particular linguistic behaviour is associated with this very cognitive function.

It is also important to note that language production as a highly complex and dynamic process is likely to be supported by a number of cognitive resources, of which inhibitory control is only one. It may be that the ability to suppress non-target information does play a role in bilingual language production, as confirmed in a number of experimental studies (e.g. Levy et al., 2007; Martin et al., 2010; Morales et al., 2011), but that it contributes minimally to the fluidity of speech. Other mental processes, such as working memory and mental set shifting may be of greater importance. Working memory may be critical in language production, not only during the planning stage but also in maintaining coherence throughout the text (e.g. Linck, Osthus, Koeth & Bunting, 2014; Martin & Slevc, 2014). Without it, speakers may not be able to screen the contents of their internal speech for potential errors or inaccuracies while speaking. Shifting, an ability to switch flexibly between tasks or mental sets, and in the context of language to divert attention from a linguistic cul de sac when the speaker has “talked herself into a corner”, may be yet another important mechanism underlying speech production. To date, these potentially relevant functions in the context of prompted extended speech have remained largely unexplored. Future models should address this gap, with the aim of determining the unique contribution of executive abilities to spoken language performance.

A third question posed by the present findings relates to performance accuracy on the Stroop task. How to explain the fact that, despite a positive correlation between reaction time and error rate on the Stroop task (r=.297, p< 0.01), reformulations and silent pauses correlated solely with inhibitory control as indexed with an error rate, but not with reaction time? As the latencies increased, there was a decrease in accuracy, suggesting that a higher frequency of errors was not simply due to a speed/accuracy trade-off. While research on inhibitory control measuring performance on a variety of cognitive tasks (e.g. Stroop task, Simon task, stop signal task, anti-saccade task) typically records response latencies and proportion of errors, the latter is seldom given adequate consideration. And yet, as can be seen in the presented analysis, accuracy on inhibitory control tasks has the potential to be used as a legitimate indicator of inhibitory control. Clearly, individuals who are more prone to errors on such tasks are less efficient at inhibiting irrelevant or distracting stimuli. Debating whether the two measures go in tandem or whether one is gained at the expense of the other is beyond the scope of the current paper, but the relation between the two variables and with corrective tendencies certainly merits a more detailed investigation.

It could also be argued that the use of inhibitory control is warranted under certain conditions, which turns the question of whether or not speakers rely on inhibitory processes into a qualitative one. As suggested by Costa, Santesteban & Ivanova (2006), the extent of the cross-language interference a speaker experiences may depend on the proficiency level and the language currently in use. It may well be that speakers at a higher level of L2 proficiency rely less on inhibitory processes and resort more to language specific selection mechanisms. This argument ties in neatly with Green's (1998) model of inhibitory control, which assumes that the mechanism is both reactive and proportional to the level of activation. In other words, inhibition in bilingual speakers is applied only after a representation from a non-intended language has been activated, and the more strongly activated the representation, the more inhibition is needed. It is possible that the highly proficient L2 speakers in the present study may not have had an opportunity to put their inhibitory capacity to use while speaking in L2 as their L1 received little or no activation.

Taken together, the present results provide insufficient evidence to support the link between inhibitory control and spoken L2 performance. Reformulations could be seen as an exception in this respect, yet a small amount of variance (ca. 5%) in their occurrence explained by inhibitory control (but only in terms of performance accuracy on the Stroop task), compounded by the fact that reformulations were treated collectively as self-repairs and false starts, does not speak in favour of such a relationship. The observation that silent pauses are more likely to occur and to be of longer duration with decreased inhibitory control can similarly be put into question based on a small variance predicted by performance accuracy on the Stroop task and the fact that this finding pertained only to silent pauses as treated collectively, without the distinction into mid-clause and end-clause silent pauses.

Limitations of the study include the fact that the Colour-Word Stroop Test is inevitably a language-based task, which is contingent on L2 proficiency. Much as it provides a measure of prepotent response inhibition, it gives an incomplete picture of the subject's language-specific inhibitory ability. The study would therefore benefit from a complementary Stroop test administered in the subject’s L1 or a non-linguistic Stroop task such as sound or picture, though these have their own culture-specific limitations. In his fifty-year review of the Stroop Color-Word Task, MacLeod (1991) reported the publication of more than 700 Stroop-related articles in the previous 25 years. He noted that, for bilinguals on the conventional colour-naming task, (1) maximal interference occurred when naming and distracting languages were one and the same (Dyer in Macleod 1991: 186), and (2) ‘If the naming language is the non-dominant one, interference between and within languages tends to be close to identical’ (Dornic, Dornic & Wirberg in MacLeod 1991: 187). This lends confidence to the application of the Stroop instrument in the present study, in which both naming and distracting languages were English, not least for consistency of administration to participants speaking 19 different first languages. Although MacLeod acknowledges the possibility of differences in orthographic and idiographic languages, he concludes that ‘the cross-language semantic contribution to Stroop interference is substantial’ (MacLeod 1991: 1987), thus justifying its use as a measure of language inhibition.

As inhibitory control is possibly one of a number of cognitive mechanisms involved in spoken language production, using a battery of tests to measure a range of executive functions would shed light on other potentially relevant processes. In addition, it would be useful to include a more comprehensive set of inhibitory control tasks, both language-specific and domain-general, as the question of whether this cognitive ability is part of the language system per se or results from generic executive control is still unresolved (e.g. Costa, 2005; de Bruin, Roelofs, Dijkstra, FitzPatrick, 2014). Adaptation of the right quantity and quality of language production tasks that lend themselves to experimental control, without being overly restrictive (e.g. a sentence production task in Engelhardt et al., 2013) and that elicit a sufficient amount and variety of disfluencies should also be on the agenda of those pursuing the link between executive functions and language production. It may be that while executive functions show individual variation, linguistic inhibition is more specific to language level, as suggested by recent studies. It therefore becomes a matter of great interest to ascertain the relationship, if any, of these two forms of cognitive inhibition and the insight it may offer into both general executive functioning and language learning, potentially opening up new areas for studies in language and cognition.

Conclusions

The goal of this study was to examine the relationship between individual variation in inhibitory capacity and the speech patterns of second language users, with a particular emphasis on speech disfluencies. The present results do not seem to support the hypothesis that the speech of individuals with poorer inhibitory capacity is characterised by reduced fluency. A series of regression analyses, in which we controlled for age and L2 proficiency, showed that the only two reliable predictions concerned silent pauses and the rate of reformulations, both in terms of self-corrections (when the speaker backtracks to correct a deviant part of an utterance) and false starts (when the speaker abandons an utterance and starts it anew). Performance accuracy on the Stroop task accounted for around 5% of variation in the rate with which utterances were reformulated and with which silent pauses were inserted into speech. Future studies could examine the contribution of inhibition and other control processes to different types of dysfluencies by employing a variety of language production tasks and gathering larger speech samples.

No relationship was found between the ability to suppress irrelevant and conflicting information and a number of speech parameters used in this analysis (repetitions, filled pauses, silent mid-clause pauses, articulation rate and performance errors). These findings indicate that inhibition may not be the most salient executive function in the service of L2 production. In future, other general processing abilities should be given more consideration to account for the individual differences with which speech is produced. The results reported in this study may also suggest that reliance on inhibitory processes in the context of prompted extended speech is contingent on the relative proficiencies of L1 and L2, where the smaller the difference between the two languages, the less the need to resort to inhibitory control.

Acknowledgement

The authors thank James Greenough, Laura Mark and two anonymous EFL teachers for their invaluable assistance with marking and the coding of variables, Matt Platts for technical assistance, and the participants for their generosity in giving their time and cooperation. They also acknowledge the helpful comments and guidance of the anonymous reviewers.

References

Abutalebi, J. & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, 20(3), 242–275.

Abutalebi, J., Annoni, J.M., Zimine, I., Pegna, A.J., Seghier, M.L., Lee-Jahnke, H., Lazeyras, F., Cappa, S.F., & Khateb, A. (2008). Language Control and Lexical Competition in Bilinguals: An Event-Related fMRI Study. *Cerebral Cortex*, 18(7), 1496-1505.

Allen, M., Powell, D. & Dolby, D. (2007). *IELTS Graduation*. Oxford, Macmillan.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., ... & Weinert, R. (1991). The HCRC map task corpus. *Language and speech*, 34(4), 351-366.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic bulletin & review,* 4(2), 221-225.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.62. Accessed on 2 October 2014 from http://www.praat.org/

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech*, 44(2), 123-147.

Bosker, H. R, Pinget., A-F., Quene, H., Sanders, T. and de Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2): 159-175

Christoffels, I.K., Firk, C., & Schiller, N.O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, 1147, 192–208.

Clark, H. H. (2006). Pauses and hesitations: Psycholinguistic approach. In Brown, K. (ed) *Encyclopedia of Language and Linguistics.* Amsterdam: Elsevier, volume V, 284-288.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73-111.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37, 201– 242.

Coderre, E. M., & van Heuven, W. J. B. (2014). Electrophysiological explorations of the bilingual advantage: evidence from a Stroop task. *PloS One*, 9(7), e103424.

Colomé, À.. (2001). Lexical Activation in Bilinguals' Speech Production: Language-Specific or Language-Independent? *Journal of Memory and Language*, 45(4), 721–736.

Colomé, À., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 96.

Corley, M. & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: the meaning of *um*. *Language and Linguistics Compass,* 2(4), 589-602.

Costa, A. (2005). Lexical access in bilingual production. In: Kroll, J.F., De Groot, A.M.B., (eds). *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, 2(03), 231-244.

Costa, A., Caramazza A., & Sebastian-Galles N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition.* 26, 1283-1296.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, 41(3), 365-397.

Costa, A. & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50(4), 491–511.

Costa A., Santesteban M., Ivanova I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32, 1057-1074.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

De Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. *NeuroImage*, 90, 348-359.

De Groot, A. M. (2011). *Bilingual Cognition: An Introduction*. Psychology Press.

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In The 6th Workshop on Disfluency in Spontaneous Speech (DiSS) (pp. 17-20).

Dell, G.S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.

DeSchepper, B. & Treisman, A. (1996). Visual Memory for Novel Shapes: Implicit Coding Without Attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1), 27-47.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication. *Studies in second language acquisition,* 20(03), 349-385.

Engelhardt, P.E., Nigg, J.T., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica*, 144(2), 424-32.

Festman, J. (2012). Language control abilities of late bilinguals. *Bilingualism: Language and Cognition*, 15(03), 580-593.

Festman, J. & Münte, T.F. (2012). Cognitive Control in Russian–German Bilinguals. *Frontiers in Psychology*, 3, 115.

Festman, J., Rodriguez-Fornells, A. & Münte, T.F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, 6(5), 1-12.

Fox Tree, J.E., & Clark, H.H. (1997). Pronouncing ‘‘the’’ as ‘‘thee’’ to signal problems in speaking. *Cognition*, 62, 151–167.

Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26(02), 349-356.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, 26(02), 275-301.

Friedman, N.P. &.Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology*, 133(1), 101-135.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production.* Vol. 1. New York: Academic Press.

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech*, 4(1), 18-26.

Gollan, T. H., & Acenas, L. R. (2004). What Is a TOT? Cognate and Translation Effects on Tip-of-the-Tongue States in Spanish-English and Tagalog-English Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 30(1), 246-269.

Green, D.W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, 27(2), 210–223.

Green, D.W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1, 67-81.

Hashemi, L., & Thomas, B. (2011). *IELTS Trainer*. Cambridge: Cambridge University Press.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of experimental psychology: Learning, memory, and cognition*, 17(1), 163.

Heidlmayr, K., Moutier, S., Hemforth, B., Courtin, C., Tanzmeister, R., & Isel, F. (2014). Successive bilingualism and executive functions: The effect of second language use on inihibitory control in a behavioural Stroop Colour Word task. *Bilingualism: Language and Cognition*, 17(3), 630-645.

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1(03), 213-229.

Hoaglin, D. C., and Iglewicz, B. (1987), Fine tuning some resistant rules for outlier labeling. *Journal of American Statistical Association*, 82, 1147-1149.

Horton, W. S., Spieler, D. H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and aging*, 25(3), 708.

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). *Aachener Aphasie Test (AAT)* Hogrefe. Göttingen, Germany.

IELTS Speaking Band Descriptors (public version). (n.d.). Available at http://www.ielts.org/pdf/UOBDs_SpeakingFinal.pdf (Accessed 24 March 2015).

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, 96(1), 29.

Kormos, J. (1999). Monitoring and Self‐Repair in L2. *Language learning*, 49(2), 303-342.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, 50(2), 343- 384.

Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145-164.

Kousaie, S., & Phillips, N. A. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain research,* 1446, 71-90.

Kroll, J.F., Bobb, S.C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition,* 9(2), 119-135.

Kroll, J. F., Bobb, S. C., Misra, M. M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes*. Acta Psychologica*, 128, 416-430.

Leal, C.F. (1995). A pause as a cause of change. ERIC Document ED 379935. Rockville, MD: Educational Resources Information Center.

Lee, M-W. & Williams, J.N. (2001). Lexical access in spoken word production by bilinguals: evidence from the semantic competitor priming paradigm. *Bilingualism: Language and Cognition*, 4(03), 233-248.

Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach. *Language Learning*, 40: 387-417.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.

Levelt, W.J.M. (1989). *Speaking*. Cambridge. MA, the MIT Press.

Levelt, W.J.M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown and P. Hagoort (Eds.), *The neurocognition of language*. Oxford, UK, Oxford University Press.

Levy, B.J., McVeigh, N.D., Marful, A., & Anderson, M.C. (2007). Inhibiting your native language: the role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, 18(1), 29-34.

Linck, J.A., Hoshino, N., & Kroll, J.F. (2008). Cross-language lexical processes and inhibitory control. *Mental Lexicon*, 3(3), 349-374.

Linck, J.A., Kroll, J.F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: evidence for the role of inhibition in second-language learning. *Psychological Science*, 12, 1507-15.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861-883.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163-203.

Martín, M. C., Macizo, P. & Bajo, T. (2010). Time course of inhibitory processes in bilingual language processing. *British Journal of Psychology*, 101(4), 679–693.

Martin, R. C., & Slevc, L. R. (2014). Language Production and Working Memory. *The Oxford Handbook of Language Production*, 437.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.

Meuter, R.F.I. & Allport, A. (1999). Bilingual Language Switching in Naming: Asymmetrical Costs of Language Selection. *Journal of Memory and Language*, 40(1), 25–40.

Misra, M., Guo, T., Bobb, S.C. & Kroll, J.F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, 67(1), 224–237.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49-100.

Morales, L., Paolieri, D. & Bajo, T. (2011). Grammatical gender inhibition in bilinguals. *Frontiers in Psychology,* 2(284), 1-9.

O'Connell, D. C., & Kowal, S. (1972). Cross-linguistic pause and rate phenomena in adults and adolescents. *Journal of Psycholinguistic Research*, 1(2), 155-164.

O'Connell, D. C., Kowal, S., & Hörmann, H. (1969). Semantic determinants of pauses. *Psychologische Forschung*, 33(1), 50-67.

O'Connell, D. C., Kowal, S., Sabin, E. J., Lamia, J. F., & Dannevik, M. (2010). Start-up rhetoric in eight speeches of Barack Obama. *Journal of Psycholinguistic Research,* 39: 393-409.

Pawley, A. & Syder, F.H. (2000). The one-clause-at-a-time hypothesis. In Riggenbach, Heidi (ed.), *Perspectives on fluency*. Ann Arbor, University of Michigan Press.

Peirce, J.W. (2006). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 15, 162(1-2), 8–13.

Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory Control and L2 Proficiency Modulate Bilingual Language Production: Evidence from Spontaneous Monologue and Dialogue Speech. *Frontiers in Psychology*, 3-57.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, 77(2), 97–132.

Poulisse, N. (1999). *Slips of the Tongue: Speech errors in first and second language production*. Amsterdam: John Benjamins B.V.

Poulisse, N. & Bongaerts, T. (1994). First Language Use in Second Language Production. *Applied Linguistics*, 15(1), 36-57.

Raupach, M. (1987). *Procedural learning in advanced learners of a foreign language* (pp. 123-155). LAUD.

Riazantseva A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, 23, 497–526.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14, 423–441.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53(2): 361-382.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662.

Tannenbaum, P. H., Williams, F., & Hillier, C. S. (1965). Word predictability in the environments of hesitations. *Journal of Verbal Learning and Verbal Behavior*, 4(2), 134-140.

Towell, R. (1987). Variability and progress in the language development of advanced learners of a foreign language. *Second language acquisition in context*, 113-127.

Van Assche, E., Duyck, W., & Gollan, T. H. (2013). Whole-language and item-specific control in bilingual language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 39(6), 1781.

Appendix A Frequencies of reported First Languages (L1)

Language	Frequency
Chinese	28
Arabic	8
Spanish	7
Thai	7
Japanese	5
Turkish	5
Kurdish	4
French	3
Russian	3
Bengali	2
German	2
Esan	1
Farsi	1
Gujerati	1
Hindi	1
Italian	1
Nepali	1
Tamil	1
Twi	1

Appendix B

Speaking prompt

Describe a journey you remember well.

You should say:

- How you travelled
- Where you went
- What happened

And explain why the journey was memorable for you.

Table 1 Descriptive statistics and Inter-rater reliability for the overall and composite L2 spoken proficiency scores

	Mean	SD	Range	ICC
Spoken L2 proficiency (overall)	5.94	.81	4.75-9	-
Fluency and Coherence	5.85	.92	4-9	.85**
Lexical Resource	6.04	.90	4.5-9	.87**
Grammatical Range and Accuracy	6.12	.84	4.5-9	.86**
Pronunciation	6.03	.95	5-9	.88**

ICC = Intraclass Correlation Co-efficient measured with a two-way, mixed model absolute agreement

** $p < .001$

Table 2 Speech parameters identified and tallied by raters in the transcribed speech samples

Parameter	Definition	Example	Comments
Filled pauses	fillers and hesitation phenomena in the form of interjections, for example, *hm.., ehr.., eh.., mm.., ehm*, and lexical fillers which introduce no semantic content within an utterance, for example, *okay.., yes.., yeah.., actually.., you know.., like..*	*I'm going to talk about the* ***her ... ehm ... eh ...*** *some study I wrote when I was working ...*	3 instances
Repetitions	unintended repeats of previously articulated material such as a phoneme, a syllable, the whole word or a cluster of words	*...because it has many* ***profe.. professional*** *singers in this programme ...*	1 instance
Reformulations	instances in which the speaker abandons an original utterance and starts it anew (false starts)	*it's better than half… than the cup is half empty …"*	1 instance – replacement of comparison with an idiom
	and instances of self-initiated corrections (self-repairs)	*it was happened ... it happened …"*	1 instance - replacement of simple past passive with simple past verb
Rater-identified Errors (proficiency and performance)	lexical	*the perception of every moment is individual in my* ***occasion***	1 instance – inappropriate use of 'occasion' for 'opinion'
	grammatical	*she* ***don't*** *use a lot of time on preparing a test*	1 instance – third person singular form required for simple present tense

Table 3 Descriptive statistics and inter-rater reliability for L2 speech measures

	Mean	SD	Range	ICC
filled pauses (%)	13.41	10.25	0.39 – 69.4	.97**
repetitions (%)	4.23	3.87	0 – 24.09	.96**
reformulations (%)	2.36	1.55	0 – 7.41	.90**
Rater-identified errors (%)	6.90	4.17	0 -18.5	.95**
performance errors (%)	1.67	1.32	0 – 7.36	-
proficiency errors (%)	5.23	3.75	0 – 15.77	-
silent pauses total frequency (%)	19.61	7.14	5.05 – 37.96	-
silent pauses total duration	24.44	9.14	7.74 – 51.28	-
silent mid-clause pauses frequency (%)	12.38	7.40	1.58 - 39.40	.96**
silent mid-clause pauses total duration	14.26	7.37	3.01 – 36.04	-
silent end-clause pauses frequency (%)	8.69	3.48	2.48 – 18.35	.68*
silent end-clause pauses total duration	12.10	6.69	3.71 – 36.99	-
words (total)	268	49.48	166 - 475	-
words (pruned)	226.51	53.2	124 - 443	-
articulation rate (words per minute)	112.15	29.31	69-216	-

*p < .05, ** p < .001

Table 4 Means (M) and standard deviations (SD) for the interference effect and the two types of trials in the Stroop and the Shape Matching Tasks expressed in RT and ER.

	RT (ms)		ER (%)	
	M	*SD*	*M*	*SD*
Stroop				
congruent	936	237	0.4	1.3
incongruent	1145	295	4.9	6
interference	209**	117	4.5**	5.9
Shape				
no distractor	991	272	1.6	3.5
distractor	1202	333	3.5	6
interference	210**	149	1.8**	6

RT = reaction time (reported in milliseconds)

ER = error rate (reported as percentage of errors)

**p<.01

Table 5 Correlations between age, spoken L2 proficiency, IC and L2 output measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. Age	1	.212	.182	.279*	.190	.123	.387**	-.008	.321**	.247*	.132	-.071	.069	-.018	-.111	.019	-.247*	-.250*	-.281*	-.266*	-.213	-.173	.083	.028	.029
Spoken L2 Proficiency Ratings																									
2. Spoken L2 proficiency (overall)		1	.944**	.917**	.927**	.899**	-.093	.010	.026	-.028	-.439**	-.412**	-.377**	-.602**	-.411**	-.523**	-.411**	-.098	-.498**	-.359**	-.396**	-.009	.480**	.617**	.658**
3. Fluency and Coherence			1	.838**	.822**	.811**	-.084	-.006	.065	-.029	-.479**	-.400**	-.371**	-.566**	-.357**	-.502**	-.488**	-.179	-.563**	-.441**	-.472**	-.069	.523**	.664**	.729**
4. Lexical Resource				1	.813**	.717**	-.001	-.053	.085	-.021	-.466**	-.391**	-.344**	-.463**	-.337**	-.395**	-.412**	-.097	-.483**	-.324**	-.417**	-.029	.482**	.619**	.635**
5. Grammatical Range and Accuracy					1	.794**	-.109	.016	.014	-.006	-.441**	-.398**	-.431**	-.659**	-.469**	-.565**	-.365**	-.057	-.441**	-.280*	-.328**	.082	.413**	.555**	.592**
6. Pronunciation						1	-.154	.088	-.074	-.047	-.225*	-.329**	-.248*	-.544**	-.362**	-.475**	-.237*	-.018	-.340**	-.271*	-.231*	-.009	.345**	.430**	.461**
Inhibitory control measures																									
7. Stroop interference (RT)							1	.297**	.206	.103	.028	.123	.150	.172	-.059	.211	-.038	-.092	-.079	-.095	-.094	-.153	.003	-.049	-.030
8. Stroop interference (ER)								1	-.015	-.092	-.075	.128	.227*	-.027	-.085	.000	.220	.231*	.042	.105	.055	.085	-.014	-.020	.002
9. Shape Matching interference (RT)									1	.053	.052	-.058	-.010	-.005	.045	-.021	-.162	-.224	-.085	-.135	-.089	-.148	-.019	-.007	.048
10. Shape Matching interference (ER)										1	.049	.107	.072	.105	-.079	.145	-.047	-.045	-.082	-.115	-.075	-.038	.126	.079	.010
Spoken L2 output measures																									
11. Filled pauses											1	.189	.165	.308**	.201	.271*	.151	-.250*	.212	-.051	.256*	-.235*	-.218	-.514**	-.540**
12. Repetitions												1	.446**	.153	.162	.113	.399**	.071	.487**	.243*	.107	-.216	-.219*	-.386**	-.389**
13. Reformulations													1	.385**	.315**	.316**	.286*	.171	.310**	.194	.083	-.167	-.228*	-.345**	-.373**
14. Raters-identified errors														1	.458**	.949**	.275*	.003	.277*	.165	.369**	-.076	-.263*	-.368**	-.435**
15. Performance errors															1	.153	.249*	.030	.303**	.146	.220	-.058	-.320**	-.359**	-.372**
16. Proficiency errors																1	.221	-.008	.201	.132	.331**	-.064	-.179	-.281*	-.351**
17. Silent pauses (total frequency)																	1	.804**	.836**	.749**	.633**	.338**	-.778**	-.741**	-.694**
18. Silent pauses (total duration)																		1	.535**	.707**	.467**	.575**	-.564**	-.399**	-.362**
19. Mid-clause pauses (frequency)																			1	.861**	.468**	.114	-.698**	-.692**	-.713**
20. Mid-clause pauses (duration)																			.	1	.310**	.243*	-.596**	-.501**	-.539**
21. End-clause pauses (frequency)																					1	.628**	-.637**	-.623**	-.582**
22. End-clause pauses (duration)																					.	1	-.426**	-.244*	-.119
23. Words (total)																							1	.929**	.832**
24. Words (pruned)																								1	.929**
25. Articulation rate																									1

***p<.05, **<.01**

Table 6 Partial correlations between inhibitory control and spoken L2 output measures after controlling for age and spoken L2 proficiency

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Stroop interference (RT)	1	.345**	.061	.003	-.146	.079	.097	.162	-.131	.223	-.012	-.004	-.011	.012	-.077	-.108	.090	.097	.095
2. Stroop interference (ER)		1	.024	-.112	-.088	.173	.274*	-.052	-.111	-.013	.220	.236*	.121	.216	.067	.115	-.040	-.049	-.022
3. Shape interference (RT)			1	-.019	-.064	-.203	-.105	.044	.107	.005	-.088	-.151	-.012	-.041	-.026	-.069	-.017	.063	.135
4. Shape interference (ER)				1	-.026	.091	.040	.063	-.105	.108	-.020	.022	-.054	-.086	-.071	.015	.169	.164	.092
5. Filled pauses					1	-.014	-.070	.049	.040	.037	.035	-.270*	.058	-.212	.176	-.218	-.022	-.351**	-.357**
6. Repetitions						1	.388**	-.116	.004	-.125	.296*	.052	.439**	.189	-.069	-.236*	-.035	-.190	-.167
7. Reformulations							1	.212	.169	.161	.203	.188	.197	.128	-.065	-.150	-.049	-.135	-.143
8. Raters-identified errors								1	.343**	.933**	.068	-.032	-.045	-.105	.192	-.098	.037	.009	-.035
9. Performance errors									1	-.019	.103	-.018	.175	.063	.063	-.077	-.156	-.158	-.165
10. Proficiency errors										1	.033	-.027	-.115	-.136	.180	-.075	.100	.070	.026
11. Silent pauses (total frequency)											1	.836**	.790**	.696**	.542**	.310**	-.735**	-.720**	-.661**
12. Silent pauses (total duration)												1	.534**	.706**	.454**	.550**	-.619**	-.483**	-.448**
13. Mid-clause pauses (frequency)													1	.821**	.311**	.031	-.640**	-.629**	-.641**
14. Mid-clause pauses (duration)														1	.156	.181	-.572**	-.449**	-.484**
15. End-clause pauses (frequency)															1	.673**	-.554**	-.550**	-.487**
16. End-clause pauses (duration)																1	-.423**	-.266*	-.153
17. Words produced (total)																	1	.916**	.809**
18. Words produced (pruned)																		1	.897**
19. Articulation rate																			1

*p<.05, **p<.01

Table 7 Linear models of variables predicting individual L2 spoken output measures

	B	SE B	β	t	*p*
Filled pauses	***R=.508, R^2=.258, adj R^2=.196, SE=9.22***				
Age	.452	.192	.289	2.355	.021*
Spoken L2 proficiency	-6.414	1.340	-.509	-4.785	.000**
Stroop interference (RT)	-.010	.011	-.115	-.959	.341
Shape interference (RT)	.000	.006	-.007	-.064	.949
Stroop interference (ER)	-.068	.187	-.039	-.362	.719
Shape interference (ER)	-.058	.179	-.034	-.323	.747
Repetitions	***R=.445, R^2=.198, adj R^2=.216, SE=3.63***				
Age	-.007	.075	-.012	-.095	.925
Spoken L2 proficiency	-1.895	.527	-.398	-3.597	.001**
Stroop interference (RT)	.002	.004	.054	.432	.667
Shape interference (RT)	-.001	.002	-.051	-.458	.648
Stroop interference (ER)	.080	.074	.122	1.084	.282
Shape interference (ER)	.067	.071	.104	.946	.347
Reformulations	***R=.470, R^2=.221, adj R^2=.189, SE=1.41***				
Age	.036	.025	.149	1.435	.156
Spoken L2 proficiency	-.797	.200	-.415	-3.983	.000**
Stroop interference (RT)	.001	.002	.066	.570	.570
Shape interference (RT)	.000	.001	-.054	-.494	.623
Stroop interference (ER)	.061	.027	.230	2.258	.027*
Shape interference (ER)	.006	.028	.023	.215	.830
Silent pauses (total frequency)	***R=.362, R^2=.131, adj R^2=.093, SE=9.92***				
Age	-.357	.178	-.231	-2.008	.049*
Spoken L2 proficiency	-.640	1.509	-.049	-.424	.673
Stroop interference (RT)	.004	.011	.041	.315	.753
Shape interference (RT)	-.007	.007	-.117	-.947	.347
Stroop interference (ER)	.459	.198	.261	2.319	.023*
Shape interference (ER)	.027	.198	.016	.135	.893
Silent pauses (total duration)	***R=.344, R^2=.118, adj R^2=.080, SE=8.8***				
Age	-.333	.157	-.245	-2.113	.038*
Spoken L2 proficiency	-.386	1.337	-.034	-.289	.774
Stroop interference (RT)	.000	.010	-.004	-.033	.974
Shape interference (RT)	-.008	.007	-.152	-1.239	.220
Stroop interference (ER)	.354	.175	.229	2.020	.047*
Shape interference (ER)	.020	.174	.014	.116	.908
Mid-clause silent pauses (frequency)	***R=.387, R^2=.150, adj R^2=.073, SE=8.8***				
Age	-.146	.130	-.135	-1.129	.263
Spoken L2 proficiency	-4.169	.906	-.479	-4.603	.000**
Stroop interference (RT)	-.005	.007	-.083	-.706	.483
Shape interference (RT)	.000	.004	-.012	-.116	.908

Stroop interference (ER)	.074	.127	.062	.587	.559
Shape interference (ER)	-.048	.121	-.041	-.400	.691
Mid-clause silent pauses (duration)	***R=.441, R^2=.194, adj R^2=.127, SE=6.8***				
Age	-.131	.143	-.117	-.916	.363
Spoken L2 proficiency	-3.108	.998	-.345	-3.113	.003**
Stroop interference (RT)	-.006	.008	-.100	-.802	.425
Shape interference (RT)	-.003	.005	-.065	-.582	.562
Stroop interference (ER)	.158	.140	.128	1.130	.262
Shape interference (ER)	-.079	.134	-.065	-.591	.556
Performance errors	***R=.443, R^2=.196, adj R^2=.130, SE=1.2***				
Age	-.131	.143	-.117	-.916	.363
Spoken L2 proficiency	-3.108	.998	-.345	-3.113	.003**
Stroop interference (RT)	-.006	.008	-.100	-.802	.425
Shape interference (RT)	-.003	.005	-.065	-.582	.562
Stroop interference (ER)	.158	.140	.128	1.130	.262
Shape interference (ER)	-.079	.134	-.065	-.591	.556
Articulation rate	***R=.678, R^2=.459, adj R^2=.414,, SE=20.5***				
Age	-.790	.427	-.194	-1.850	.068
Spoken L2 proficiency	23.207	2.984	.706	7.777	.000**
Stroop interference (RT)	.022	.023	.097	.953	.344
Shape interference (RT)	.010	.014	.068	.741	.461
Stroop interference (ER)	-.109	.417	-.024	-.262	.794
Shape interference (ER)	.296	.400	.067	.740	.462

*p<.05, **p<.01

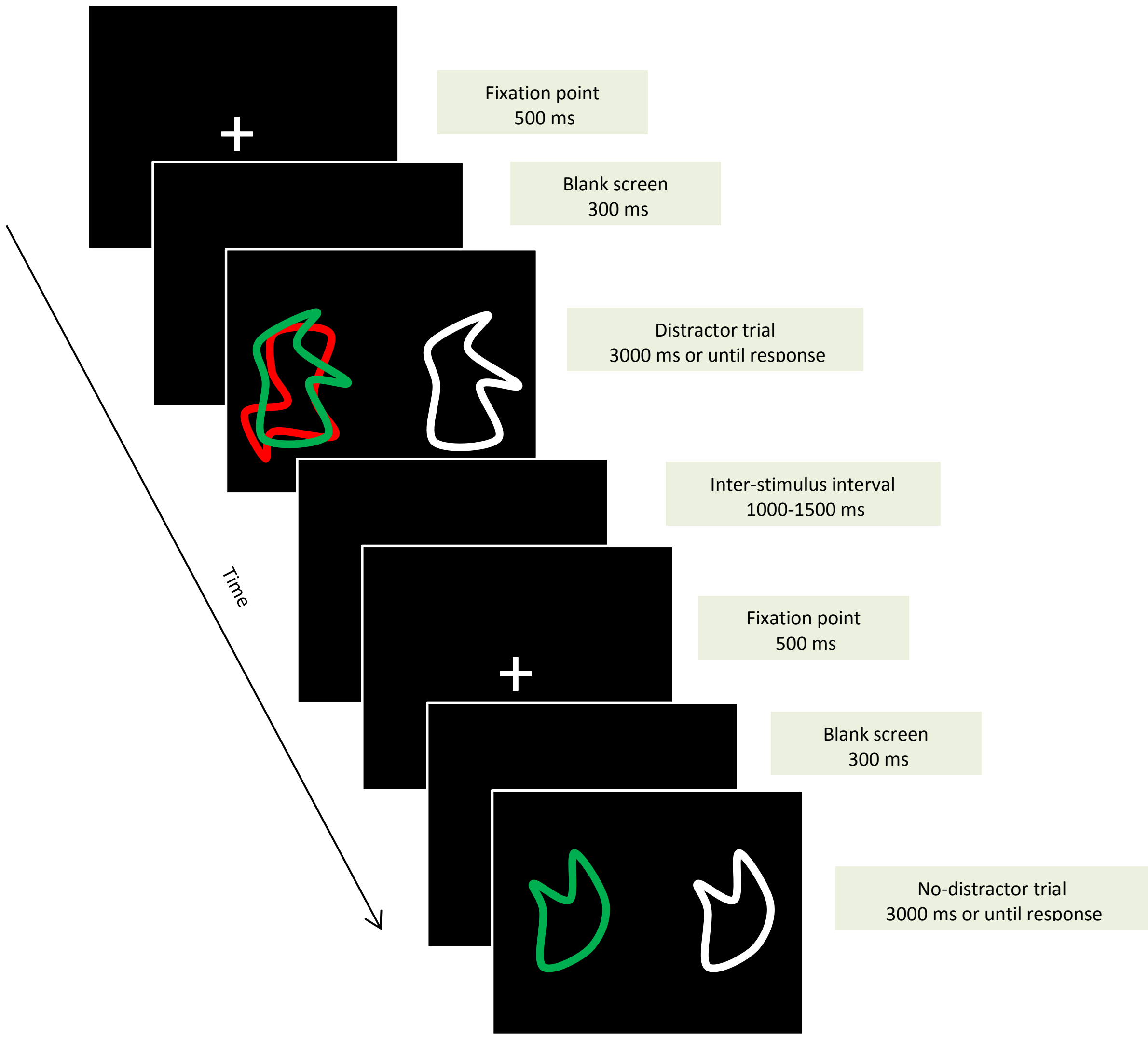

Figure 1 Shape Matching Task. Participants indicated whether the green target shape on the left matched the white shape on the right, ignoring the red distractor shape when one was present.